R E P O R T R E S U M E S

ED 003 444 24

COURSES OF MISRECALL OVER LONG-TERM RETENTION INTERVALS AS RELATED TO STRENGTH OF PRE-EXPERIMENTAL HABITS OF WORD ASSOCIATION.

BY- BILODEAU, EDWARD A. BLICK, KENNETH A.
TULANE UNIV. OF LOUISIANA, NEW ORLEANS

REPORT NUMBER CRP-S-030 PUB DATE 65
EDRS PRICE MF-$0.09 HC-$0.96 24P. PSYCHOLOGICAL REPORTS, MONO. SUPPL. 6-V16,1965,1173-1192PP.

DESCRIPTORS- *LEARNING, LEARNING EXPERIENCE, LEARNING PROCESSES, ASSOCIATIVE LEARNING, TIME FACTORS (LEARNING), *LEARNING THEORIES, *WORD RECOGNITION, *STIMULUS BEHAVIOR, *STIMULUS DEVICES, RETENTION STUDIES, NEW ORLEANS, LOUISIANA, KEESLER AIRFORCE BASE

THIS STUDY WAS MADE TO COMPARE THE EFFECTS OF STIMULATION AND NONSTIMULATION ON RECALL OF WORDS FOLLOWING TIME-DELAY PERIODS. THE SUBJECTS (670 AIRMEN) WERE TRAINED WITH AN EXAMPLE WORD LIST AND TWO WORD LISTS CONTAINING FIVE OF THE SECONDARY WORDS ASSOCIATED WITH RUSSELL-JENKINS STIMULUS WORDS. AFTER TIME DELAYS OF 2 MINUTES, 20 MINUTES, 2 DAYS, OR 28 DAYS, 24 GROUPS OF SUBJECTS WERE TESTED ON RECALL OF THE SECOND LIST. ON THE RECALL TEST, HALF OF THE SUBJECTS WERE GIVEN THE FIVE STIMULUS WORDS CORRESPONDING TO THE SECOND WORD LIST, AND HALF WERE NOT. THOSE SUBJECTS TESTED AFTER 28-DAY TIME DELAYS STILL SHOWED STRONG EVIDENCE OF THE WORD LIST TRAINING, BUT THE PERFORMANCE WAS MORE LIKE THAT OF FREE ASSOCIATING SUBJECTS THAN THAT OF SHORTER RETENTION SUBJECTS. IN THE NONSTIMULATED GROUP, INTRUSIONS FROM UNIDENTIFIED SOURCES WERE MORE NUMEROUS WHEN THE TIME PERIOD WAS LONGER. AS RECALLS OF WORDS ON LIST TWO DECREASED WITH TIME, INTRUSIONS OF WORDS FROM THE EXAMPLE LIST AND FROM THE FIRST TRAINING LIST WERE MORE NUMEROUS WHERE CULTURAL PROBABILITIES SUGGESTED THIS OUGHT TO HAPPEN. (AL)

S-030

PSYCHOLOGICAL REPORTS

Monograph Supplement 6-V16 1965 price 1.00

Not sent to Library of Congress Photoduplication Service for microfilming ALR

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

COURSES OF MISRECALL OVER LONG-TERM RETENTION INTERVALS AS RELATED TO STRENGTH OF PRE-EXPERIMENTAL HABITS OF WORD ASSOCIATION

EDWARD A. BILODEAU
Tulane University

AND

KENNETH A. BLICK
Randolph-Macon College

ED003444

PSYCHOLOGICAL REPORTS

Psychological Reports is published bimonthly, two volumes a year, the first with issues in February, April, and June, and the second with issues in August, October, and December, from Box 1441, Missoula, Montana. The subscription rate is $30.00 per year, single issues are $8.00. Subscriptions are accepted only for full calendar years.

The purpose of this journal is to encourage scientific originality and creativity, in the field of general psychology, for the person who is first a psychologist, then a specialist. It carries experimental, theoretical, and speculative articles; comments; special reviews; and a listing of new books and other material received. Controversial material of scientific merit is welcomed. An attempt is made to balance critical editing by specific suggestions as to changes to meet standards. Articles should be submitted in duplicate to one of the editors or to Box 1441, Missoula, Montana. In the former case, a copy of the covering letter should be sent to Box 1441, Missoula, Montana.

Publication is in order of receipt of proof from the authors. Authors pay for reprints, including costs of cuts and special composition. There are three publication arrangements:

(1) *Regular articles.* These are articles which print to from 2 through 20 pages. Authors receive 200 reprints. Reprint charge is $15.00 per page in multiples of four pages. Additional reprints can be ordered.

(2) *One-page notes and articles.* This arrangement is useful where the author does not intend to do further work in an area but feels that preliminary findings should be put on record, where it is expected that it will be several years before the final study is completed and reported, where the author submits a summary of a study accompanied by the full report for filing with the American Documentation Institute, or where a particular finding can be reported completely in one page. Authors receive 50 reprints. Reprint charge is $15.00. Additional reprints can be ordered.

(3) *Monograph supplements.* Some papers printing to more than 20 pages are published as monograph supplements. These are distributed to subscribers of *Psychological Reports* as part of regular issues, and are also made available as separates. Authors receive 200 reprints of the monograph with covers. Reprint charge is $15.00 per page in multiples of four pages. Additional reprints can be ordered.

Articles should be prepared carefully, following the form suggested for publications of the American Psychological Association (see *Psychol. Bull.*, 1952, 49, 389-449, the 1957, 1958, or 1965 revision). Either the old or the new APA form will be acceptable during 1965 and 1966. Proof is mailed to the author within approximately three weeks of acceptance of his article, and reprints are mailed within five weeks of return of proof by the author.

It is the policy of this journal to file raw data with the American Documentation Institute wherever possible. Authors should submit appropriate tables with their articles. In this way, a permanent file of data available to all researchers can be built up.

PRECEDING PAGE(S) MISSING

Psychological Reports, 1965, 16, 1173-1192. © Southern Universities Press 1965
Monograph Supplement 6-V16

COURSES OF MISRECALL OVER LONG-TERM RETENTION INTERVALS AS RELATED TO STRENGTH OF PRE-EXPERIMENTAL HABITS OF WORD ASSOCIATION[1]

EDWARD A. BILODEAU[2] AND KENNETH A. BLICK
Tulane University *Randolph-Macon College*

CONTENTS

Method 1178
Results 1180
Discussion 1185
References 1191

Summary.—Following a proaction paradigm, 670 *Ss* were trained with two lists containing five of the secondary associates (R_2) to Russell-Jenkins stimulus words, and were tested 2 min., 20 min., 2 days, or 28 days later for retention of the second list. During recall, half of the *Ss* were administered the five stimulus words corresponding to the five R_2 words of the second list (stimulated condition), and half did not receive the stimulus words (not-stimulated condition). The stimulus words were divisible into three categories, in effect varying the cultural associative probabilities at each retention interval. Altogether, there were 24 groups completing a $4 \times 2 \times 3$ factorial design. The retention of R_2s decreased with time and the effect of stimulation was to raise their level of production above groups not so stimulated. As R_2s decreased with time, intrusions of R_1s (primaries) and R_{3-n}s (sum of R_3 to R_n) became more numerous where the cultural probabilities suggested this ought to happen. After 28 days, *Ss* still showed strong evidence of the training exposure but performance was more like that of free-associating *Ss* than that of shorter retention groups. In the not-stimulated condition, intrusions from unidentified sources (classified as Remainder) were more numerous the longer the retention interval. Collectively, these data support the conclusion that the amount of proactive interference via specific pre-experimental word-word habits increases as a function of time. An item analysis suggested a monotonic pattern of rs for forgetting under stimulated conditions, but not under conditions of free recall. This was interpreted to mean that forgetting when stimulated was more a process of complication than simplification and resembles a process sometimes found in motor-skills retention. Other correlational analyses proved useful tools for describing forgetting; questions pertaining to the behavior of items were quite as intriguing as those about *Ss*. Also, more variance could be accounted for after long than short retention intervals.

Bilodeau, Fox, and Blick (1963) have demonstrated the effects of cultural word habits on the kinds of correct recall and misrecall to be expected after train-

[1]This work was supported by the Small Grant Program of the U. S. Office of Education. We wish to thank Paul W. Fox and David C. Howell for their assistance and suggestions. To M. B. Jones of the University of Florida we owe special gratitude for his critical reading of the section on *item analysis*.
[2]The writers wish to acknowledge the assistance of the Commander and Colonel E. E. Cockrum of Keesler AFB in providing so many *Ss* for a difficult temporal schedule.

ing on a short list of words. The essence of the procedure involved the use of a stimulus during the recall period to assist in the retrieving process. The recall stimulus was a word associated in various degree to the word presented during training, but nevertheless a word which had not been administered to *S* during training. The stimulus word used by *E* was chosen because of its particular properties of response evocation exhibited in tests of free association. The free-association test had been administered earlier, elsewhere, and to other people. Hence, the findings showed *cultural* characteristics of associative structures of words to be a means of modifying and predicting the probability of a correct retrieval or the probability of a particular misrecall.

A few examples from free-association data below will make the above clear. These examples have been selected from Categories A, B, and C, classifications of associative structure described by Bilodeau, Fox, and Blick (1963).

Category C

	S_1	R_1	R_2	R_{3-n}
	light	dark	bulb	bright, etc.
p		.57	.06	.27

Dark is the primary association to *light* with $p(R_1) = .57$, *bulb* is the secondary association to *light* with $p(R_2) = .06$, and *bright, etc.* are the remaining associates in the response hierarchy and their combined probability $[p(R_{3-n})]$ equals .27. If *S* were trained on *bulb* and tested for recall of *bulb* in the presence of *light*, we might expect that (a) *bulb* would appear more frequently than when *light* is not used, (b) when *bulb* is not recalled, *dark* might intrude or appear in its stead, and (c) if not *dark* intruding, then *bright, etc.* A Cat. C word is characterized as a stimulus where $p(R_1)$ is maximum in the normative hierarchy.

Category B

	S_1	R_1	R_2	R_{3-n}
	ocean	water	sea	blue, etc.
p		.31	.30	.39

In another case, R_1 and R_{3-n} form less competition to recalling R_2 because $p(R_2)$ is at its maximum, and consequently R_2 acts as a stronger competitor to R_1 and R_{3-n}. Here, R_2 should be easier to recall than in the previous example. If R_2 is not recalled, then the intrusion is more likely to be an R_{3-n} and less likely to be an R_1 than for words from Cat. C. A Cat. B word is characterized as a stimulus word where $p(R_2)$ is maximum, although still less than $p(R_1)$ in the cultural norms.

Category A

	S_1	R_1	R_2	R_{3-n}
	cheese	cracker	milk	cake, etc.
p		.11	.10	.79

For Cat. A words, the stimulus word ought to perform differently again. The $p(R_2)$ and $p(R_1)$ are low, while $p(R_{3-n})$ is high. In this case, R_2 should be recalled about as well as in C above, but intrusions which would predominate are the R_{3-n} variety. A Cat. A word is characterized as a stimulus word where $p(R_{3-n})$ is maximum in the cultural norms.

The Russell-Jenkins (1954) free-association norms were analyzed as suggested by the examples of categories above and 15 words were found which satisfied each of the three response patterns. The categories were defined as follows:

Cat. A: Low $p(R_1)$, low $p(R_2)$, and high $p(R_{3-n})$
Cat. B: Moderate $p(R_1)$, maximum $p(R_2)$, and moderate $p(R_{3-n})$
Cat. C: High $p(R_1)$, low $p(R_2)$, and low $p(R_{3-n})$

In the first experiment on retention (Bilodeau, Fox, & Blick, 1963), it was found that the pattern of obtained response recalls matched the associative structure of the free-association norms.

In 1959, Deese suggested that free-association data could be used as a basis for understanding memory processes. Until quite recently, however, the relationship between word-association data and verbal-habit interference has been neglected. The method reported by Bilodeau and his associates for predicting interference from pre-experimental associations has been joined by those developed by Coleman (1963) and Turnage (1963). With response latency as an index of associative strength, Coleman made use of the individual's own free associations in constructing paired-associate lists designed to produce high and low interference. He found that high-interference lists (the response to be learned was a weak associate) were more difficult to learn and retain, and in addition, the majority of intrusions occurring during learning and recall were traced to previous stronger associations. One variable employed by Turnage in paired-associate learning was the differential associative strength of the response terms to the stimulus terms. Retention was assessed by the method of recognition after 90 sec. and 7 days. By an analysis of errors during recall, associative errors were found to increase over time and their incidence was positively related to their associative strength. Turnage concluded that long-term forgetting is due to the recovery of extinguished pre-experimental response associates which compete with the correct response. There appears to be a rising trend in the number of studies using strength of pre-existing associations as the major predictor of extra-experimental interference. Bilodeau, Levy, and Sulzer (1963) have published a monograph on the various types and uses of motor and verbal stimulation and have shown that until recently research in learning has emphasized the acquisition of behavior under a wide variety of training conditions and shied away from retention behavior and wide varieties of recalling conditions. The associative network of probabilities among words now appears to be the most promising of all recall variables.

Fox, Blick, and Bilodeau (1964) later simplified their experimental procedure in a second study of verbal stimulation and the original findings were confirmed. The present study is patterned after theirs, which should be consulted for details, except that four values of the retention interval are used instead of the one 2-min. interval previously employed. The present study uses 2 min., 20 min.,

2 days, and 28 days as retention intervals and was designed to show how pre-experimental language habits become more and more involved with what is retrieved the longer the interval.

Previously, several investigators have examined the temporal course of interference for verbal habits, and the general conclusion is that interference, both proactive and retroactive, increases as a function of time (Briggs, 1954; Briggs, Thompson, & Brogden, 1954; Postman, 1962; Underwood, 1948; Underwood, 1949). The present approach represents a substantial departure from these studies in that the assessment of pre-experimental verbal habits by cultural word-association data is the major predictor of the types of intrusions to be expected with increases in time. For our purposes this variable supplants frequency of usage which has been employed on several occasions by Postman (1961, 1962, 1964) and was an integral part of the original theoretical analysis of extra-experimental sources of interference (Underwood & Postman, 1960).

Control of Recall Environment

The category of stimulation takes on special significance with retention interval because of hypothesized interactions. Administering Cat. B stimuli should have the effect of retarding forgetting whereas A and C should enhance the probabilities of intrusions of R_{3-n}s and R_1s, respectively. The effect of training *S* on R_2 words should be to increase their probability of occurrence markedly over the short term along with a corresponding decrease in the frequency of R_1s and R_{3-n}s as compared with free-associating *S*s. In the long run, however, with the decay of $p(R_2)$, there might well be a rising trend in R_1s and R_{3-n}s as the *S*s revert toward behaving as free-associating *S*s. *S*s having been trained with R_2s, however, might never again be able to produce norm-like data exactly, but it is hypothesized that the consequence of increasing time is a movement toward normative behavior.

The normative probability values calculated earlier (Bilodeau, Fox, & Blick,

TABLE 1

NORMATIVE RESPONSE PROBABILITIES CLASSIFIED BY CATEGORY OF STIMULATION AND POSITION IN RESPONSE HIERARCHY

Category		Position in Response Hierarchy		
		$p(R_1)$	$p(R_2)$	$p(R_{3-n})$
A	R-J*	.14	.12	.74
	F-B†	.16	.10	.74
B	R-J	.32	.24	.44
	F-B	.35	.21	.44
C	R-J	.73	.05	.22
	F-B	.67	.06	.27

*Calculated from data of Russell and Jenkins (1954).
†Calculated from data of Fox and Bilodeau (1963).

1963) for the Russell-Jenkins free-association norms are reproduced in Table 1 for inspection. [Cats. A, B, and C are there defined by the obtained values of $p(R_1)$, $p(R_2)$, and $p(R_{3-n})$.] The second entries are norms obtained more recently from airmen at Keesler AFB. Since the experiment to be reported was also done at the Technical Training School, Keesler AFB, was decided that the general reliability of the assignment of words to categories should be ascertained. If we were to use the Minnesota Norms of Russell-Jenkins, it would mean using norms from college men and women obtained in 1953 as a basis for predicting responses of airmen in a technical school on a military base 10 years later. This was actually what happened in the two earlier studies mentioned before. It was decided that it would be safer to possess two sets of norms (Minnesota and Keesler). Since norms were to be collected, it was also thought desirable to improve upon the methodology of free-association testing. So in comparing the results in Table 1, it should be kept in mind that the norms are not replications for they differ both in subject population and procedure. Nevertheless, Table 1 shows a considerable level of agreement between the means from norm to norm. The average deviation for the nine comparisons is close to 2 percentage points. The Keesler and Minnesota norms, however, do not have all their R_1s and R_2s in common. Values in Table 1 use the Minnesota words as a basis. Table 2 shows the degree of commonality between the primary and secondary words of the two schools. As expected, the trends are markedly related to the category, i.e., as $p(R_1)$ rises so does agreement in the identification of the primary term. For Cat. B, in Minnesota, $p(R_2)$ was a maximum, meaning that R_2 though in second place was the strongest competitor to R_1. Reversals in Cat. B were expected and indeed in 15 opportunities, 4 Minnesota R_2s became R_1s at Keesler. Cat. A is the least reliable with few R_2s in common (.13) and worthless for producing both R_1 and R_2 in common (.00). The lack of agreement in Cat. A is intimately related to its definition. It deals with words of low probability of evocation and the failure to identify words-in-position in common is a manifestation of the basic property of Cat. A.

Though we see from Table 2 that words may shift about in rank and lower the word commonality index, Table 1 has shown that the mean strengths of association, estimated in the two places under different procedures, are comparable.

TABLE 2
COMMONALITIES BETWEEN KEESLER AND MINNESOTA NORMS

	Stimulus Category*		
	A	B	C
Primaries in Common	.47	.80	1.00
Secondaries in Common	.13	.40	.47
Both in Common	.00	.40	.47

*Base of 15 words.

In the remainder of this paper Minnesota words and Keesler AFB probabilities for these words will be used exclusively. The procedures for collecting these norms are described in the section below. They are especially important details because they were designed to be more closely related to methodologies of retention testing than were those of Russell-Jenkins. After all, if one is to use free-association data as a basis of understanding retention data, both should rest on similar technical bases wherever standardization and reconciliation are possible.

Fox-Bilodeau Free-association Procedures

In July of 1963 45 stimulus words from the Kent-Rosanoff Word Association Test were administered to 107 airmen students in the School of Electronics Principles, Keesler AFB, Biloxi, Mississippi in a test of discrete free association. The airmen were comparable in age and education to college freshmen and were tested in small classes of 15 to 25 *S*s. The stimulus words were selected according to their pattern of associative probabilities previously described as Cats. A, B, and C (Bilodeau, Fox, & Blick, 1963).

To the customary procedures of collecting free-association norms, several special modifications were introduced. These include: (a) reduction of *S*'s word load from 100 to 45 stimulus words; (b) presentation of 15 stimuli per page (one category per page), and control of between page effects through a 3 × 3 Latin square schedule of categories; (c) control of word-order effects through the systematic variation of word orders within pages. Procedures (a), (b), and (c) are modeled after those used in retention testing in the earlier studies. Still, there is a difference in that a 5- to 10-word load (not a 45) is typical of the retention studies. In the ideal norm, *S* has free-associated once to one word. In the limiting case, the norms would not be biased by experimental proactions and serial orders.

Method

Subjects and Apparatus

Over 700 airmen from Keesler AFB were unsystematically assigned to 24

TABLE 3
Experimental Design and Number of *S*s Per Group

Category of Stimulation	Retention Interval			
	2 Min.	20 Min.	2 Day	28 Day
Stimulated				
A	26	28	27	30
B	24	29	28	29
C	24	30	28	30
Not-stimulated				
A	30	27	25	30
B	30	27	26	31
C	30	26	24	31

groups. A total of 670 of these were available for the recall session, their distribution across experimental treatments being given in Table 3.

A booklet suitable for training and testing large intact groups was used. The basic components have been described elsewhere (Bilodeau, Fox, & Blick, 1963). The booklet consists of an opaque cover sheet, training page 1, training page 2, vowel cancellation page, and recall page. The cover sheet contains a line for *S*'s name, the first and second pages of training exhibit five R_2 words each, the cancellation page contains a large number of vowels and consonants, and the recall page shows five blanks where the stimulus words might also be shown. Every page is headed by a few words of written instruction.

Procedure

The design was a $4 \times 2 \times 3$ factorial as shown in Table 3. There were 4 values of retention interval (2 min., 20 min., 2 days, and 28 days), two of stimulation (stimulated and not-stimulated), called S and NS, respectively, and three categories of stimulation (Cat. A, Cat. B, and Cat. C).

Every booklet took the proaction form of training on one list, training on a second list, and testing for recall of the second list. During training, *S*s first studied five R_2 words from one category and then five R_2 words from another category. For example, when Cat. A words were assigned to the second page, Cat. B or C words were used on the first training page, but this variation is not at issue in this monograph. A category contains 15 stimulus words which were randomly divided into three lists of five R_2 words. The three lists within categories averaged 9 *S*s.

For the stimulated groups the recall page contained the five stimulus words corresponding to the five R_2 words on the second page and five empty lines for answers. The stimulus words were absent for NS.

*S*s were allowed 20 sec. per training page where the instructions at the top read: "Study the words shown below." After the appropriate retention interval, *S*s were asked for recall of the R_2 words from the second training page. The recall sequence actually began with a period of 60 sec. of vowel cancellation before the recall test. After the cancellation period, the instructions for the stimulated groups read: "A while ago you studied some words on two pages, ——— and ———. The words below may help you remember those words. Write down the words that appeared on the second page, page ———. You must fill in every line. If you cannot remember, write in the first word that the printed word makes you think of."[3] *S*s were allowed 120 sec. for recall and, as indicated above, were instructed to free-associate if they could not remember the

[3]The first and second training pages were labeled A and B, respectively, and the As were hand-printed in black ink and the Bs in red. This color difference was continued in the recall instructions where one A and two Bs appeared. The letter-labeling and color-coding were attempts to preclude the confusion of page membership through a misunderstanding or an inattention to the twoness of the lists.

training words. The printed instructions for NS were essentially the same except that the reference to "the words below" was omitted. The printed instructions were preceded by brief verbal instructions from *E* who reminded the *S*s that they had been trained minutes, days, or weeks ago and were going to participate in a "similar experiment."

*S*s were trained and tested in intact groups of about 90 (these 90 contained *S*s of the three categories). The eight retention treatments (4 retention intervals $\times$ 2 of stimulation) were scheduled for training on different days and all were tested for recall on the last day of the experiment. This procedure reduced the possibility of *S* interaction on the all-important day of recall. Our best assurance against much gossip about the experiment was the fact that *S* had to report immediately to his flight and continue his daily regimen after a recall session.

Booklets and pencils were distributed by four assistant *E*s who served as proctors. To reduce the possibility of copying a neighbor's work, the booklets were distributed so that only every ninth booklet was identical. All page-turning was paced by *E* who directed *S*s when to turn and stop with every page. In addition, *S*s were told to follow the instructions printed at the top of the page.

Additional Details

Several precautions were taken in an effort to hold constant from one retention interval to another the most important activities which occurred between training and recalling. The 2-min. and 20-min. groups were trained and tested in *one* session while the 2-day and 28-day groups required *two* sessions and two days. In order to make the retention-interval activities of the shorter-term groups more comparable to those of the longer-term groups, a disengagement procedure was effected in the short-term groups by having them get up and change seats immediately after training (see Boswell & Bilodeau, 1964). In addition, as stated before, all groups were required to cancel vowels for 60 sec. immediately preceding the test for recall. Thus all four groups were disengaged immediately after training and were reengaged immediately before recalling. The advantage of such a procedure is to provide control over some important intervening activities of all four treatments, thereby making it possible to place their data along a similar continuum.

The 3350 (670 $\times$ 5) recall responses were classified according to one of seven sources of recall. The seven, in order of precedence, included position in the Russell-Jenkins hierarchy (R_1, R_2, R_{3-n}), Same List (a correct word written in the wrong place on the recall page), Other List (an R_2 intrusion from the first training page), Remainder (none of the previous five), and Blank (where *S* failed to respond).

RESULTS

The probability of correct words (R_2s) recalled in the correct position is

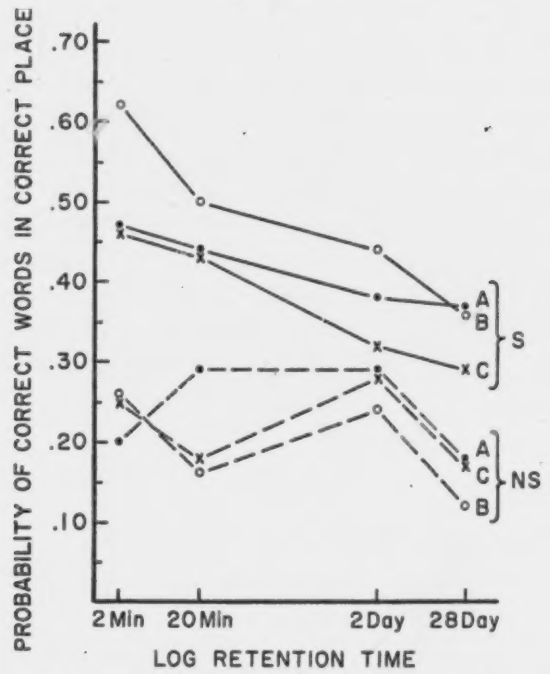

FIG. 1. The probability of correct words in the correct place as a function of log retention time for Categories A, B, and C under Stimulated (S) and Not-stimulated (NS) conditions

plotted in Fig. 1.[4] The duration of the retention period serves as the independent variable and S and NS as the parameter.[5] First, every stimulated group produces more R_2s than its corresponding not-stimulated group.[6] Second, the R_2s of the stimulated conditions drop significantly with increasing retention time ($F = 5.22, df = 3/321, p < .01$) whereas the effect of retention intervals fails to achieve the .05 level ($F = 2.33, df = 3/325$) for the not-stimulated conditions. For a comparison of the means of NS at 28 days versus 2 min., however, the F was significant at the .05 level ($F = 4.13, df = 1/180$). Third, there is a significant between-category effect within-S ($F = 3.09, df = 2/321, p < .05$), Cat. B words being recalled best over three intervals (B > A + C, over-all at .05 level), but there is no significant Category effect for the not-stimulated con-

[4]A set of four tables of data has been deposited with the American Documentation Institute. Copies may be obtained by citing Document No. 8411 and by remitting $1.25 for photoprints or 35-mm. microfilm. Make checks or money orders payable to: Chief, ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C.

[5]The retention interval ranges from 2 to over 40,000 min. and is most easily expressed with a log scale.

[6]Since Group 2 Min. almost duplicates one in the study by Bilodeau, Fox, and Blick, many of the present results are established phenomena, reducing the need for statistical tests. The present analysis stresses trends across increasing retention intervals. For comparisons of NS versus S and comparisons among categories of the 2-min. point, the reader is referred to the earlier study.

ditions ($F < 1.00, df = 2/325$). Fourth, the trends in Fig. 1 are such that there is no obvious way of summarizing what might happen after 28 days of retention.

Many R_2 words of training are retrieved and written in the wrong position, these being called Same-list intrusions. An analysis *irrespective of position* implies summing Same-list intrusions with the data of Fig. 1. This analysis produced a marked reduction in the difference between S and NS, leaving Cat. B, Stimulated as the best of the three categories eliciting R_2s. Since this category is precisely the one which was expected to produce the greatest effect, the grand mean difference $p(B_S - B_{NS})$ of .10 was evaluated ($F = 6.81, df = 1/216$) and is significant beyond the .01 level. This is taken as evidence that the stimulus words not only serve as cues to assign responses from line to line but also serve to arouse dormant correct responses.

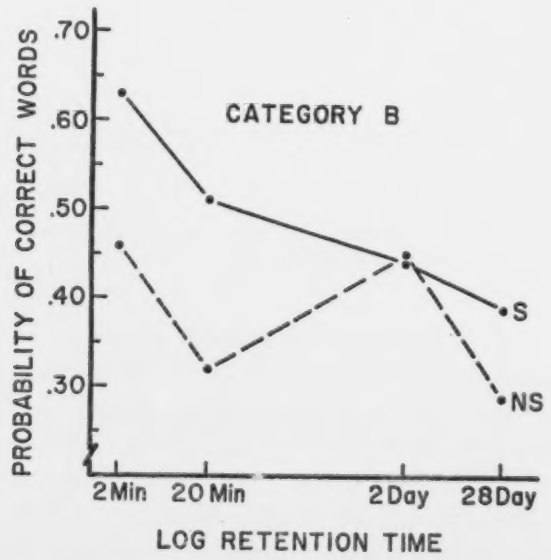

FIG. 2. The probability of correct words (irrespective of position) in Category B as a function of log retention time for Stimulated (S) and Not-stimulated (NS) conditions

The mean combined probability of R_2 words for Cat. B at each level of retention is shown in Fig. 2. At the 2-day point, NS inexplicably exceeds S by a whisker, otherwise the data look as expected. Despite the unfortunate appearance of the not-stimulated point at 2 days, the over-all trend is downward ($F = 5.76, df = 3/110, p < .05$). A design with anything less than 28 days of retention might well have led to null results with respect to the production of R_2s. Of course, forgetting might still be manifest in differences among categories or in intrusions of R_1s or R_{3-n}s. The R_2 data for Cats. A and C are not graphed because the stimulated treatments should not exceed NS and they did not. Rather,

when stimulated, Cats. A and C should produce extra numbers of intrusions of R_1s and R_{3-n}s, analyses which follow.

The intrusions of R_1s during retention are pictured in Fig. 3. The largest effect is between categories where all subgroups are appropriately ranked with

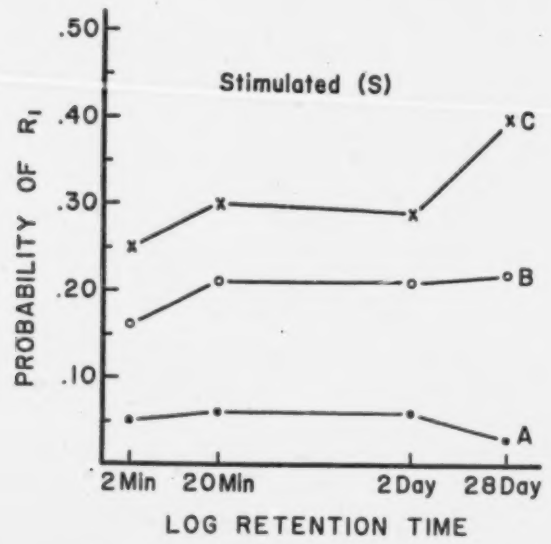

FIG. 3. The probability of intrusion of R_1 as a function of log retention time for Categories A, B, and C under Stimulated conditions

C > B > A at each level of retention ($F = 38.75$, $df = 2/330$, $p < .01$). At 28 days, the differences between C, B, and A are greatest, the major change in rate of intrusion occurring in C which is producing about 40% R_1s. Control *S*s, who free-associate only, produce 67% R_1s. An analysis of Cats. A and B yielded non-significant *F* ratios and indicates that the occurrence of R_1 intrusions in Cats. A and B does not increase with retention time. However, Cat. C, which is characterized by high $p(R_1)$, does show a significant rise in R_1 intrusions between 2 min. and 28 days. The latter effect was tested by a one-tailed *t* test which was significant at the 5% level ($t = 1.73$, $df = 52$). Cat. A produces few R_1s even in free association, the mean in free association being 16%. Data for not-stimulated groups averaged $p = .01$ intrusions for R_1, a figure so small as to discourage graph making.

The meaning of Fig. 3 is clear. To produce intrusions of R_1s, three variables should be considered: (a) presence or absence of stimulus, (b) stimulus category, and (c) retention interval. The results of Cat. C show that at the 28-day point of retention, *S*s are producing more R_1s than when tested 2 min. after training and they are becoming more like free-associating *S*s though they are still not producing R_1s with normative frequency.

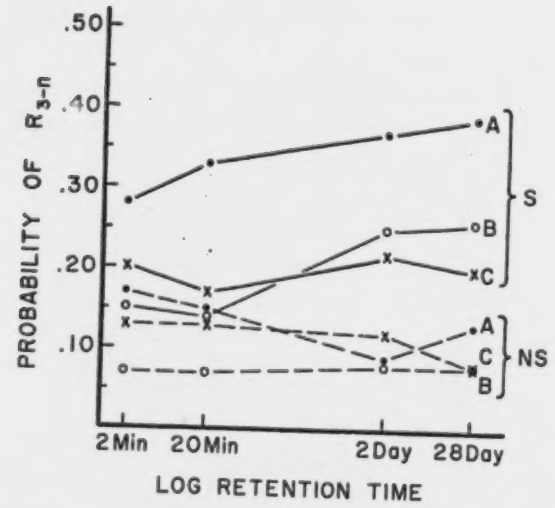

FIG. 4. The probability of intrusion of R_{3-n} as a function of log retention time for Categories A, B, and C under Stimulated (S) and Not-stimulated (NS) conditions

The production of R_{3-n}s is depicted in Fig. 4. The not-stimulated conditions produce a few (about 10%), but the amount is independent of retention time. The differences between categories also fail to attract attention. When stimulation is used, however, the trend for all categories is upward with Cat. A in the lead because Cat. A represents those stimulus words which maximize R_{3-n} under free-association techniques. Cat. B was the only stimulated group to ex-

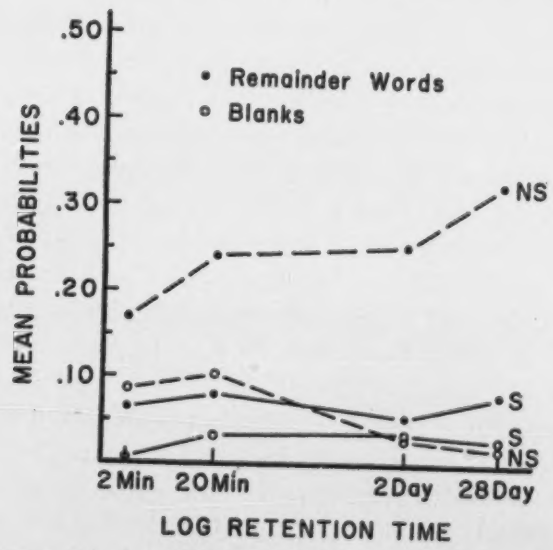

FIG. 5. The probability of Remainder words and Blanks as a function of log retention time for Stimulated (S) and Not-stimulated (NS) conditions. Each data point represents the sum of three categories of stimulus words.

hibit a significant rise in R_{3-n}s ($F = 3.00$, $df = 3/106$, $p < .05$). Also, when the stimulus is provided during recall, there is a significant between-category effect ($F = 14.88$, $df = 2/330$, $p < .01$). The R_{3-n} words occur 74% of the time in free-association norms of Cat. A, a value which might be considered as the limit for Cat. A after the complete elimination of a training residual.

Fig. 5 shows what happens to Remainder words and Blanks as a function of the retention interval. The probabilities were averaged over categories because there were such small differences among them. The results in Fig. 5 show the S condition produced few Remainders and Blanks and also show that the Remainders for NS rise with increasing retention interval until they constitute over 30% of the data at Day 28 ($F = 6.40$, $df = 3/333$, $p < .01$). With the exception of Remainders for NS, Remainders and Blanks do not constitute important sources of data under the conditions of this experiment. The same is true of Other-list intrusions as a source of data; the stimulated groups run about 5% or less and not-stimulated groups about 15% with no trend with retention interval.

Discussion

In general, the retention functions have risen, fallen, and remained constant depending upon the type of response measure, use of stimulation or no stimulation, category of stimulation, and the value of the retention interval. The classical curve of retention plots correct responses retrieved as a function of duration of retention interval and this curve falls. The correct response in the right place (R_2) fell too, but its decay with time was retarded by stimulation. The apparent effect of stimulation was to heighten the probability of retrieval of R_2s over the entire 28-day period. Stimulation generally served to assign R_2 to the proper line of the test page. It did not serve to elicit a greater rate of correct responses over-all except for the critical Cat. B.

Assuming 100% recall at 0 min., the form of the retention function for both types of correct responses (see Figs. 1 and 2) from 0 min. to 28 days takes the classical shape of sharp negative acceleration (Ebbinghaus, 1885; Luh, 1922). There is, however, no good reason to assume 100% recall at 0 min. After 2 min., the course of forgetting can be described, though conservatively, by a negative, linear function with no more forgetting occurring at one interval than another.

There was a noticeable difference in R_2s and R_{3-n}s between Cats. A, B, and C *only* when stimulus words were administered (see Figs. 1 and 4). The three categories of words fail to maintain their associative identities when the stimulus words are omitted during recall. Thus, it is concluded that effects of category hold only under conditions of stimulation.

As the probability of correct responses decreased with time, the probability of certain other intruding sources was found to increase. After the initial training session with R_2s, interference from extra-experimental sources should be expected to increase the longer one waits for the recall of the R_2s. Intrusions of

R_1s and R_{3-n}s were numerous in stimulated groups, rising over time and depending on the category of stimulus words. Both R_1s and R_{3-n}s were found to progressively interfere with the recall of R_2s. Therefore, this experiment did satisfy the purpose of showing the temporal course of interference by extra-experimental word-word habits.

The above findings support the work of other investigators (Briggs, 1954; Briggs, Thompson, & Brogden, 1954; Postman, 1962; Underwood, 1948; Underwood, 1949) who have shown that proactive interference of verbal habits increases over time as second list responses fade and first list or extra-list responses become more frequent. The type of forgetting in the present study is attributable to the progressive recovery of pre-existing word-word probabilities, and the data provide convincing evidence that proactive interference from extra-experimental sources increases as a function of time.

Other-list intrusions (R_2s but from the wrong page) and Blanks occurred but tended to be relatively infrequent and without trend. The comparative unimportance of Other-list intrusions, which is an often-studied source of interference in many experiments, was not unexpected. The words of the first page bore no relation that we can discern to those of the second page. These data support an earlier contention (Bilodeau, Fox, & Blick, 1963) and recent evidence that a knowledge of the strength of words related through probability networks represents a good strategy for predicting the kind and amount of interference between lists. A relatively high level of interference was found by Blick (1965) who tested for the stimulated recall of R_2s when the corresponding R_1s occupied the other training list. The occurrence of R_1s during recall was directly related to their cultural associative strengths, and for one group the intrusion rate was as high as 55%, a rate far exceeding ordinary expectations.

Remainder words occurred more often when stimulation was not used and an increasing frequency with time was found. The Remainder classification contained responses which (a) could not be found in the cultural hierarchy, (b) were not intrusions from the Same or Other list, or (c) were not failures to respond. As such, Remainder represents a source of unaccountable variance in misrecall and is larger in free recall than stimulated recall. Remainder words become quantitatively more important over time until they constitute 30% of the recall data at 28 days.

Finally, it should be stated that particular differences and trends in this experiment are nowhere very large in absolute magnitude. It must be noted that *seven* response measures are being used over a 100-point data space (from p of 0.00 to 1.00) and act to restrict the magnitude of change over time. For example, given $p = .50$ for responses recalled correctly, the remaining .50 must be shared by six other measures of recall. Strong movements in one source of variance preclude action in others.

Item Analysis

Stimulation should change not only the mean recall as discussed above but also the position of the item (word) among the distribution of words. A word which is easily retrieved after 2 min. may or may not be comparatively easy to retrieve after a month. This section deals with the phenomena of consistencies in item retrieval across the four values of retention. The questions that are immediately raised are many and the way is uncharted. See the work of Runquist (1957) for a start on this objective.

After 2 min. in free recall, the word *child* as shown below, ranked rather poorly among the 45 words. At 28 days it was retrieved with about the same success (35th vs 34th rank).

	2 Min.	20 Min.	2 Day	28 Day
NS Rank	34	40	18	35

Under conditions of stimulated recall the magnitude and trends for *child* are different, the ranks being shown below. *Child* was selected to make the point be-

	2 Min.	20 Min.	2 Day	28 Day
S Rank	33	34	3	9

cause it makes it rather well: the stimulation became effective with the two longer intervals. Even in absolute terms the comparison is much the same. The chances of retrieving *child* under stimulation at 2 min., are only 8 percentage points over NS, while at 28 days, S is superior to NS by 28 percentage points.

Table 4 shows how five words of Cat. B behave under S and NS treatments, Cat. B words being chosen because the facilitation by stimulation is expected to

TABLE 4

FREQUENCY OF RETRIEVAL OF R_2 WORDS FROM CATEGORY B FOR STIMULATED (S) AND NOT-STIMULATED (NS) TREATMENTS USING THE BEST 5 WORDS OF 15

R_2 Words	2 Min.		20 Min.		2 Day		28 Day	
	NS	S	NS	S	NS	S	NS	S
Water	4	3	2	5	2	3	0	7
Sheep	6	7	3	6	4	6	5	4
Scared	5	7	0	9	7	6	2	9
Sea	7	8	2	7	4	9	5	7
Red	6	5	1	6	2	8	2	5
Σf	28	30	8	33	19	32	14	32
p^*	.38	.39	.19	.45	.33	.53	.32	.54

*Proportion (p) of words correct contributed by 5 of 15 words from Category B.

be favorable. Of the 15 Cat. B words available for examination, only the five (top third) producing the highest total evocation of R_2 (summed over four retention intervals) were selected. Table 4 shows that these 5 words produce more than one-third of the recalls. As a matter of fact, they also produce a rising proportion of recalls the longer the interval. For comparison, under NS conditions these same 5 words produce about one-third of the recalls regardless of the value of the retention interval. It is clear that stimulation can lead to a greater amount of recall and also a repositioning of items within the distribution. The Pearson correlation coefficient will be used below to express the changes in internal consistency from interval to interval.

The proportion of correct recalls (including those written in the wrong place by *S*) for each of the 45 R_2 words was calculated for treatments S and NS. The values of p ranged from 1.00 to nearly 0.00 for both S and NS at 2 min., meaning that a few words were recalled on every occasion and a few at the other extreme were almost never retrieved. The wide range of difficulty among items naturally facilitated the analysis. At 28 days, the range was only somewhat smaller, though it was quite obvious by inspection that words especially well (or poorly) retrieved at 2 min. were not necessarily well (or poorly) retrieved again at 28 days. *Dirt*, for example, was poorly recalled at 2 min. under treatment NS ($p = .10$, Rank = 44.5), but after 28 days, it was better recalled on both absolute and relative bases ($p = .50$, Rank = 11.5). Word frequency counts were examined (Thorndike & Lorge, 1944) to determine whether this variable could account for differential memorability among the training words. It was readily apparent from the start that our stimuli and secondaries were common in the language and that frequency of use would show little difference among the words. Thus, it did not seem plausible that the great range of obtained variability among words could be explained by frequency of use. Instead, this might be done by referring to differences in associative networks, an hypothesis now being checked in the Tulane laboratory.

Some inconsistency across retention intervals is due to error or unreliability, yet some of it may be real and our purpose is to see if a portion of it can be related to the conditions of recall. Because of recency of exposure to two lists, it is not implausible that competition between words of the two lists may be keener early in retention than late. If the retrieval potential of a particular word were to improve over the retention interval without the use of stimulation, then it would probably be harder to improve upon this by stimulation than for the word which tends to move in the opposite direction. Also, under stimulated conditions, competition (from R_1 and R_{3-n}) as shown earlier, is not constant over retention and an item could be expected to change its position as stimulation became effective. This change might have a pattern.

The matrix for NS is shown at the top of Table 5, that for S is shown at the bottom. There is no striking thing about the matrix for NS except for the uni-

TABLE 5
INTERCORRELATION MATRICES FOR ITEM ANALYSES OF $p(R_2)$ UNDER TWO CONDITIONS OF RECALL

Condition of Stimulation	Retention Interval		
	20 Min.	2 Day	28 Day
Not-stimulated			
2 Min.	.58	.55	.62
20 Min.		.56	.65
2 Day			.59
Stimulated			
2 Min.	.51	.40	.46
20 Min.		.42	.44
2 Day			.30

formity of the entries which run to a few points to either side of $r = .59$. There is no trend over retention and variates spanning longer intervals of time intercorrelate as well as (sometimes better than) those spanning smaller intervals. What this means is not clear. It does mean no comfort for those who would expect the highest correlation at the upper left and the smallest at the upper right, where because of error or forgetting, variates spanning greater time intervals ought to correlate less well than for shorter intervals. Instead, the matrix for NS shows that in the absence of experimental stimulation the items quickly find their levels and remain stable.

The matrix for S is different from NS in that (a) all the *r*s are lower (the highest of S being lower than the lowest of NS) and (b) there is a trend or pattern in the matrix. The pattern places the highest value in the upper left, and the lowest in the lower right. As a matter of fact, the pattern is the monotonic hierarchy described by Jones (1962) and previously discussed by Bilodeau, Jones, and Levy (1964) for its implications for a theory of retention. The latter investigators found the same pattern of *r* in matrices of motor-skills data. The curious thing about monotonic hierarchies is that variates spanning greater intervals intercorrelate stronger than those spanning lesser times as, for example, 2 min. $\times$ 28 days $>$ 2 days $\times$ 28 days. The monotonic pattern cannot be interpreted as produced by a process of simplification. Hence, there is no evidence here that forgetting is a process of simplification or unlearning. Rather it seems to be a process involving progressive complications. (The belief that learning is a process of simplification over trials is wide-spread; see Bilodeau & Bilodeau, 1961.) Bilodeau, Jones, and Levy (1964) have rationalized the monotonic pattern for motor-retention data by suggesting the accumulation of random variance. In the present study the source of this variance change is clearer to us than it was to them, for it can be traced to the various stimuli introduced by *E* into the retention-test environment. Perhaps, but not certainly, if *S* had been used in all four re-

tention groups (repeated recalling), the monotonic pattern would have been exaggerated for there would be four instead of one source of interference. The present study differs from the earlier in that independent groups of *Ss* are used at each interval. Items, not *Ss*, are repeated at each retention interval. The monotonic pattern cannot, then, be solely attributable to repeated retrieving and accumulating random variance. But whatever the explanation, the finding of another monotonic structure is a signal event.

The effect of the stimulation upon the retrieval of R_2 words is shown in two ways. First, it lowers the inter-session reliability in the test-retest sense. That is, the discrimination between words is better ascertained under NS than S. The effect of using stimulation at 2 min. and 20 min., for example, is to produce inconsistency from one distribution to the other. Here we are comparing the absolute magnitudes of the coefficients in the two portions of Table 5. Second, stimulation changes the pattern within the matrix by introducing sources of variance whose effect depends upon the time the retention stimulation is introduced. The consequence of introducing the stimulus at 28 days is different from and more serious than its introduction at 2 min. and so it cannot be said that an R_2 is under constant conditions of competition from other words.

TABLE 6

CORRELATIONS FOR ITEM ANALYSES OF $p(R_2)$ UNDER STIMULATED (S) AND NOT-STIMULATED (NS) CONDITIONS OF RECALL

	Retention Interval			
	2 Min.	20 Min.	2 Day	28 Day
S × NS	.42	.00	.10	−.09

Table 6 shows the *r*s when S and NS are used as the variates. At the 2-min. interval *r* is .42, showing rather little variance in common between the variates in short-term retention, i.e., the difficulty of the words is largely a matter of the condition of retrieving. This *r* of .42 is even less than that obtained by correlating within NS for groups as much as 28 days apart as in Table 5. At 20 min., 2 days, and 28 days in Table 6 there is probably no communality at all between the effects of S and NS. This is taken to mean again that the two treatments interact with the retention interval in different ways.

TABLE 7

INTERCORRELATION MATRIX FOR ITEM ANALYSIS OF $p(R_1)$ FOR STIMULATED RECALL

Interval	Retention Interval		
	20 Min.	2 Day	28 Day
2 Min.	.55	.54	.70
20 Min.		.65	.60
2 Day			.66

Table 7 is the matrix for items involving the intrusion of primaries [$p(R_1)$] as the datum. Since NS produced too few R_1s for such an analysis, only the one for S was computed. Inspection of Table 7 shows no obvious trend, and in this respect and in absolute magnitude as well, matches the matrix of R_2s for NS. The alignments among items under conditions of stimulation cohere more for R_1 than for R_2 and the reason is not obvious. It may be that the intrusion of R_1 is a stronger event as measured by item analysis than by mean probability as in Fig. 3 discussed earlier. To say that there is no trend to Table 7 is conservative and to say that the upper-right entry is largest is not. Jones (1960) has named the latter phenomenon an *apical* hierarchy, and not having retention and error trends in mind, has interpreted the apical hierarchy as structurally degenerate with no existing intelligible psychological interpretation.

TABLE 8
INTERCORRELATION MATRIX FOR ITEM ANALYSIS OF $p(R_1)$, $p(R_2)$, AND $p(R_{3-n})$ FOR STIMULATED RECALL

	Retention Interval			
	2 Min.	20 Min.	2 Day	28 Day
$p(R_2) \cdot p(R_1)$	−.13	−.34	−.28	−.48
$p(R_2) \cdot p(R_{3-n})$	−.58	−.56	−.58	−.41
$p(R_1) \cdot p(R_{3-n})$	−.28	−.36	−.42	−.34

Finally, the relationships of R_2 to R_1 and R_{3-n} can be investigated as in Table 8. The rs are all negative as expected, showing that words which produce more of the one produce less of the other as the variates compete for the data space. It should be recalled that the sum of the variates [$p(R_2) + p(R_1) + p(R_{3-n})$] will ordinarily vary about .86. The competition between R_2 and R_1 seems to be greatest after the 28-day interval where an r of −.48 is obtained. All of the variance at 28 days can be accounted for by referring to R_1, R_2, R_{3-n} and Remainder simultaneously, but even more important is that a substantial piece of the 28-day variance can be explained by referring to the coefficients of previous retention intervals as in Table 5. If a theory of forgetting suggests that less and less variance is accounted for with increasing retention interval, these data do not support it.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. MCD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, 12, 243-280.

BILODEAU, E. A., FOX, P. W., & BLICK, K. A. Stimulated verbal recall and analysis of sources of recall. *J. verb. Learn. verb. Behav.*, 1963, 2, 422-428.

BILODEAU, E. A., JONES, M. B., & LEVY, C. M. Long-term memory as a function of retention time and repeated recalling. *J. exp. Psychol.*, 1964, 67, 303-309.

BILODEAU, E. A., LEVY, C. M., & SULZER, J. L. Long-term retention under conditions of artificially induced recall of related events. *Percept. mot. Skills*, 1963, 16, 895-910. (Monogr. Suppl. 8-V16)

BLICK, K. A. Cultural primaries as a source of interference in short-term verbal retention. *J. exp. Psychol.*, 1965, 69, 246-250.

BOSWELL, J. J., & BILODEAU, E. A. Short-term retention of a simple motor task as a function of interpolated activity. *Percept. mot. Skills*, 1964, 18, 227-230.

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, 47, 285-293.

BRIGGS, G. E., THOMPSON, R. F., & BROGDEN, W. J. Retention functions in reproductive inhibition. *J. exp. Psychol.*, 1954, 48, 419-423.

COLEMAN, E. B. The association hierarchy as an indicator of extra-experimental interference. *J. verb. Learn. verb. Behav.*, 1963, 2, 417-421.

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, 58, 17-22.

EBBINGHAUS, H. *Ueber das Gedaechtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker & Humbolt, 1885. (Trans. as *Memory: a contribution to experimental psychology* by H. A. Ruger and C. E. Bussenius. New York: Teachers College, Columbia Univer., 1913.)

FOX, P. W., BLICK, K. A., & BILODEAU, E. A. Stimulation and prediction of verbal recall and misrecall. *J. exp. Psychol.*, 1964, 68, 321-322.

JONES, M. B. Molar correlational analysis. *USN Sch. Aviat. Med. Monogr.*, 1960, No. 4.

JONES, M. B. Practice as a process of simplification. *Psychol. Rev.*, 1962, 69, 274-294.

LUH, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, 31, No. 3 (Whole No. 142).

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, 61, 97-110.

POSTMAN, L. The temporal course of proactive inhibition for serial lists. *J. exp. Psychol.*, 1962, 63, 361-369.

POSTMAN, L. Acquisition and retention of consistent associative responses. *J. exp. Psychol.*, 1964, 67, 183-190.

RUNQUIST, W. N. Retention of verbal associates as a function of strength. *J. exp. Psychol.*, 1957, 54, 369-375.

RUSSELL, W. A., & JENKINS, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test.* Contract No. N8onr-66216, Office of Naval Research and Univer. of Minnesota, 1954. (Tech. Rep. No. 11)

THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words.* New York: Teachers College, Columbia Univer., 1944.

TURNAGE, T. W. Pre-experimental associative probability as a determinant of retention. *J. verb. Learn. verb. Behav.*, 1963, 2, 352-360.

UNDERWOOD, B. J. "Spontaneous recovery" of verbal associations. *J. exp. Psychol.*, 1948, 38, 429-439.

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, 39, 24-34.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, 67, 73-95.

Accepted May 3, 1965.